STATISTICAL ANALYSIS PLAN

Building Positive Relationships with your Teen: Evaluating the Teen Triple P Programme

The University of Warwick

Principal investigators: Kylie Gray and Paul Thompson

YOUTH ENDOWMENT FUND

Building Positive Relationships with your Teen: Evaluating the Teen Triple P Programme

Statistical analysis plan

Evaluating institution: The University of Warwick
Principal investigator(s): Kylie Gray and Paul Thompson
V1.2 10.07.2022

YEF statistical analysis plan

Project title[1]	Building Positive Relationships with your Teen: Evaluating the Teen Triple P Programme
Developer (Institution)	Triple P UK
Evaluator (Institution)	CEDAR, University of Warwick
Principal investigator(s)	Professor Kylie Gray, Dr Paul Thompson
SAP author(s)	Dr Paul Thompson, Professor Kylie Gray, Professor Peter Langdon, and Professor Richard Hastings
Trial design	Two arm cluster randomised controlled trial with internal pilot
Trial type	Efficacy with internal pilot

Evaluation setting	Family / local authority
Target group	Families (parents/carers) of 11 to 15 year olds at the edge of care
Number of participants	412 parents/carers (approx. 275 families), and up to 412 adolescents
Primary outcome and data source	*6 months post-randomisation SDQ externalising score – Parent report*
Secondary outcome and data source	(1) Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999) parent report internalising problems scale, adolescent report externalising and internalising problems scales. (2) Parenting practices: Parenting Scale Adolescent version (PSA) parent report (3) Adolescent prosocial behaviours – The Prosocial behaviour subscale of the SDQ (Goodman, 1999). Adolescent and parent report. (4) Adolescent peer relationships – The Peer Relationship Problem subscale of the SDQ (Goodman, 1999). Adolescent and parent report. (5) Interparental outcome - Dyadic Adjustment Scale (DAS-7) is a 7-item measure (Sharpley & Cross 182; Hunsley et al., 2001). Parent report. (6) Parent mental health – the Kessler 6 (Kessler et al., 2003. Parent report. (7) Parent Wellbeing - The Warwick-Edinburgh Mental Well-being Scale (SWEMWBS; Tennant et al., 2007). Parent report. (8) Conflict Behavior Questionnaire (CBQ-20) (Robin & Foster, 1989). Parent and adolescent report. (9) Child-parent relationship – the Closeness subscale of the Child-Parent Relationship Scale (CPRS, short form) measures (Pianta, 1995). Parent report. (10) Family functioning – the Family APGAR scale. (Adapatability, Partnership, Growth, Affection and Resolve; APGAR; Smilkstein, 1978). Parent report.

	(11) Parent self-regulation – the Parenting Self-Regulation Scale is a 12-item parent-completed measure of parental-regulation (Tellegen et al., 2022). (12) Antisocial behaviours - Self Report Delinquency Scale (SRDS; Smith & McVie, 2003)
Potential moderators	1. Out of home placement 2. Learning disabilities (LD): Estimated Verbal IQ based on two subtest of the Wechsler Abbreviated Scale of Intelligence. 3. Ethnicity
Planned number of sites	6 local authorities across England
Inclusion criteria	Families of young people aged 11-15 years determined as being on the *edge of care* <u>Definition of *Edge of Care*</u> *Edge of Care* refers to children/young people who either: • have not entered into care as they have been assessed and the LA has chosen to support them and their families through alternative provisions/services. Or • they are being considered for care but have not entered into local authority care.
Exclusion criteria	• Families where one or more parent has received a multi-session parenting programme covering similar content to Triple P over the previous 12 months • Families where one or more parent is currently receiving a multi-session parenting programme covering similar content to Triple P or any multi-component manualised family intervention, such as Multi-Systemic Therapy

Treatment duration	10 sessions over 10 weeks
Follow-up duration	6 months and 12 months post-randomisation
Planned trial period	40 months
Primary objective	Determine whether there is a benefit of support as usual (SAU) plus Standard Teen Triple P (Teen Triple P) over SAU alone improving parent/carer rated adolescent externalising behaviour problems at 6-months post-randomisation in adolescents at the edge of care.
Secondary objectives	1. Complete an Internal Pilot in the first year to inform the decision as to whether proceeding with a definitive trial is warranted and feasible. 2. Determine whether TEEN TRIPLE P + SAU, (a) reduces parent reported adolescent internalising behaviour and increases prosocial behaviours at 6 months and 12 months post-randomisation, and (b) reduces adolescent reported externalising and internalising behaviour problems and increases prosocial behaviours at 6 and 12 months post-randomisation, (c) improves parenting practices, parent self-regulation, interparental relationships, parental-adolescent relationships and parental well-being at 6 and 12 months post-randomisation, decreases parent reported antisocial behaviours at 12 months post-randomisation, and (d) reduces the chance of a child going into out of home placement over a 12 month period. 3. Carry out exploratory sub-group analyses of outcomes by adolescent learning disability status and whether living with foster versus biological/adoptive parents. 4. Monitor and report and adverse events related to TEEN TRIPLE P. 5. Complete a process evaluation using key indicators drawn from the logic model, including an evaluation of acceptability

	and the experiences of parents/carers, adolescents with a broad range of ethnic and diverse backgrounds, and other key stakeholders (e.g., practitioners, delivery team), and fidelity of delivery of TEEN TRIPLE P.
Intervention	Standard Teen Triple P (TEEN TRIPLE P). The programme involves parents attending 10 (1-hour) one-to-one sessions, where they learn practical strategies for how to manage their child's problematic behaviour, promote healthy development, and improve the quality of the parent-child relationship. Sessions are delivered face-to-face, either in person or via video-conferencing. All parents will be invited to attend sessions.

SAP version history

Version	Date	Changes made and reason for revision
1.0	25-05-2023	Initial draft
1.1	10/07/2023	Addressing YEF reviewer comments. Added pilot sample size justification, additional effect size justification from protocol, and minor edit to add an alternative analysis. Other typos and minor edits also made.
1.2	15/08/2023	Addressing comments by YEF internal review.

ROLES AND RESPONSIBILITIES

Trial Statistician:	Miss Atiyya Nisar		
Role:	Research Fellow/Trial Manager		
Date:		Signature:	
Senior Statistician:	Dr Paul Thompson		
Role:	Assistant Professor in Applied Statistics / senior trial statistician		
Date:		Signature:	
Chief Investigator (s):	Professor Kylie Gray / Dr Paul Thompson		
Role:	Professor / Assistant Professor (CEDAR, University of Warwick)		
Date:		Signature:	
Other non-signatory contributor to the SAP:		Professor Richard Hastings	
Role:	Head of Department, Professor, and Cerebra Chair of Family Research (CEDAR, University of Warwick)		

Table of contents

Introduction

This statistical analysis plan provides guidelines for the final presentation and analysis for the Building Positive Relationships with your Teen: Evaluating the Teen Triple P Programme (Teen TRIPLE P trial). This plan, along with all other documents relating to the analysis of this trial, will be stored in the Statistical Analysis Master File electronically and/or in hard signed copy formats.

Background

Rationale and research question

Many UK families with young people at the edge of care experience multiple and long-standing difficulties, including mental ill-health, violence, substance misuse, and relationship and behavioural difficulties (Ofsted, 2011). Young people are more at risk of entering the out-of-home care system when experiencing social disadvantage, maltreatment, parental substance misuse, or maternal depression (Simkiss et al, 2013; NICE Guideline, No.26. 2015). Drivers of adolescent out of home placements are associated with family stress and breakdown, and adolescent behavioural problems (Percy-Smith et al, 2018).

An intervention is needed to address these risk factors and reduce care placements, thus changing the trajectory for young people and their families. Evidence-based parenting intervention strategies supported by social care services can support families at the edge of care (Bezeczky et al 2020; National Council of Voluntary Child Care Organisations, 2007; Ofsted, 2011).

Evidence-based interventions that address a number of risk and protective factors for the development of youth behaviour and emotional problems, as well as violence and delinquency in adolescence and adulthood, can improve social, behavioural and emotional outcomes for adolescents, enhance positive parenting practices, reduce family conflict, and reduce disruptive teenager behaviour (Wetherall, 2010; Salari et al., 2014).

By improving parenting skills and the parent-child relationship, overall family functioning and adolescent emotional and behavioural adjustment improves. A key focus of all Triple P interventions is to train parents to generalise the parenting skills developed throughout the program to new problems, situations and to all relevant siblings.

Objectives

Primary objective (PO)

Determine whether there is a benefit of support as usual (SAU) plus Standard Teen Triple P (TEEN TRIPLE P) over support as usual (SAU) in improving parent/carer rated adolescent externalising behaviour problems at 6-months post-randomisation in adolescents at the edge of care.

Secondary objectives (SO)

1. Complete an Internal Pilot in the first year to inform the decision as to whether proceeding with a definitive trial is warranted and feasible.

2. Generate evidence to consider whether TEEN TRIPLE P + SAU, (a) reduces parent reported adolescent internalising behaviour and increases prosocial behaviours at 6 months and 12 months post-randomisation, and (b) reduces adolescent reported externalising and internalising behaviour problems and increases prosocial behaviours at 6 and 12 months post-randomisation, (c) improves parenting practices, parent self-regulation, interparental relationships, parental-adolescent relationships and parental well-being at 6 and 12 months post-randomisation, decreases adolescent parent reported antisocial behaviours at 12 months post-randomisation, and (d) reduces the chance of a child going into out of home placement over a 12 month period.

3. Carry out exploratory sub-group analyses of outcomes by adolescent learning disability status and whether living with foster versus biological/adoptive parents.

4. Assess the sensitivity of findings under different assumptions with respect to missing data.

5. Monitor and report and adverse events related to TEEN TRIPLE P.

6. Complete a process evaluation using key indicators drawn from the logic model, including an evaluation of acceptability and the experiences of parents/carers, adolescents with a broad range of ethnic and diverse backgrounds, and other key stakeholders (e.g., practitioners, delivery team), and fidelity of delivery of TEEN TRIPLE P.

7. Monitor and report adverse events related to TEEN TRIPLE P.

Study Materials

Trial design

A two-arm cluster (with families as the clusters) randomised efficacy RCT of Standard Teen Triple P (TEEN TRIPLE P) plus SAU vs. SAU alone, running over 40 months, involving parents and foster carers of young people aged 11-15 years on the edge-of-care. There will be 6- and 12-month follow-ups, and a process evaluation and internal pilot. 412 Parent/Carer and their Young person(s) at the edge of care will be recruited. Families will be randomised on a 1:1 basis to either intervention or control arm using stratified permuted block randomisation, stratifying by local authority. Clustering by family unit accounts for the dependency between the observations from the same family (i.e. both parents can respond as both parents can take part in the study if they choose). This approach has been routinely used in practice (Coulman et al., 2020; Bell et al. 2008).

Trial design, including number of arms		*Two-arm cluster randomised control trial*
Unit of randomisation		*Families*
Stratification variables (if applicable)		*Local authority*
Primary outcome	variable	Adolescent externalising behaviour problems
	measure (instrument, scale, source)	Parent completed Strengths and Difficulties Questionnaire (SDQ) – externalising scale (SDQ; Goodman, 1999) at 6 months
Secondary outcome(s)	variable(s)	(1) Parent reported - Adolescent behavioural and emotional problems (2) Parenting practices (3) Parent and Adolescent reported prosocial behaviours (4) Adolescent reported peer relationships (5) Interparental outcome (6) Parent mental health (7) Parent Wellbeing (8) Conflict Behavior

		(9) Child-parent relationship (10) Family functioning (11) Parent self-regulation – the Parenting Self-Regulation (12) Out of home placement (13) Adolescent antisocial behaviours
	measure(s) (instrument, scale, source)	(1) Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999) parent report internalising problems scale, adolescent report externalising and internalising problems scales at 6 and 12 months. (2) Parenting practices: Parenting Scale Adolescent version (PSA) parent report (3) Adolescent prosocial behaviours – The Prosocial behaviour subscale of the SDQ (Goodman, 1999). Adolescents and parents report at 6 and 12 months. (4) Adolescent peer relationships – The Peer Relationship Problem subscale of the SDQ (Goodman, 1999) (5) Interparental outcome - Dyadic Adjustment Scale (DAS-7) is a 7-item measure (Sharpley & Cross 182; Hunsley et al., 2001) (6) Parent mental health – the Kessler 6 (Kessler et al., 2003 (7) Parent Wellbeing - The Warwick-Edinburgh Mental Well-being Scale (SWEMWBS; Tennant et al., 2007 (8) Conflict Behavior Questionnaire (CBQ-20) (Robin & Foster, 1989) (9) Child-parent relationship – the Closeness subscale of the Child-Parent Relationship Scale (CPRS, short form) measures (Pianta, 1995) (10) Family functioning – the Family APGAR scale (Adapatability, Partnership, Growth, Affection and Resolve; APGAR; Smilkstein, 1978). (11) Parent self-regulation – the Parenting Self-Regulation Scale is a 12-item parent-completed measure of parental-regulation (Tellegen et al., 2022).

		(12) Out of home placement (13) Antisocial behaviours - Self Report Delinquency Measure (SRDM; Smith & McVie, 2003)
Baseline for primary outcome	**variable**	Adolescent externalising behaviour problems
	measure (instrument, scale, source)	Parent completed Strengths and Difficulties Questionnaire (SDQ) – externalising scale (SDQ; Goodman, 1999) at baseline
Baseline for secondary outcome	**variable**	(1) Parent reported - Adolescent behavioural and emotional problems (2) Parenting practices (3) Parent and Adolescent reported prosocial behaviours (4) Adolescent reported peer relationships (5) Interparental outcome (6) Parent mental health (7) Parent Wellbeing (8) Conflict Behavior (9) Child-parent relationship (10) Family functioning (11) Parent self-regulation – the Parenting Self-Regulation (12) Out of home placement • Adolescent antisocial behaviours
	measure (instrument, scale, source)	(1) Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999) parent report internalising problems scale, adolescent report externalising and internalising problems scales at 6 and 12 months. (2) Parenting practices: Parenting Scale Adolescent version (PSA) parent report (3) Adolescent prosocial behaviours – The Prosocial behaviour subscale of the SDQ (Goodman, 1999). Adolescents and parents report at 6 and 12 months.

		(4) Adolescent peer relationships – The Peer Relationship Problem subscale of the SDQ (Goodman, 1999) (5) Interparental outcome - Dyadic Adjustment Scale (DAS-7) is a 7-item measure (Sharpley & Cross 182; Hunsley et al., 2001) (6) Parent mental health – the Kessler 6 (Kessler et al., 2003 (7) Parent Wellbeing - The Warwick-Edinburgh Mental Well-being Scale (SWEMWBS; Tennant et al., 2007 (8) Conflict Behavior Questionnaire (CBQ-20) (Robin & Foster, 1989) (9) Child-parent relationship – the Closeness subscale of the Child-Parent Relationship Scale (CPRS, short form) measures (Pianta, 1995) (10) Family functioning – the Family APGAR scale (Adapatability, Partnership, Growth, Affection and Resolve; APGAR; Smilkstein, 1978). (11) Parent self-regulation – the Parenting Self-Regulation Scale is a 12-item parent-completed measure of parental-regulation (Tellegen et al., 2022). (12) Out of home placement (13) Adolescent antisocial behaviours - Self Report Delinquency Measure (SRDM; Smith & McVie, 2003)

All participants (in both trial arms) will complete assessments at baseline (prior to randomisation), 6- and 12-months post-randomisation and will be given a choice of how these are completed. These assessments can be completed in a number of ways, face-to-face, online on a website, via telephone, via videoconferencing, or on paper via the post. Participants, parents/guardians, young people and key stakeholders will be invited to participate in semi-structured interviews to ascertain acceptability and their experience of

the parenting programme and being in the trial. We will interview a sample of those randomised to each arm of the trial.

Randomisation

Families will be randomised with equal probability to either the intervention or comparator arm using random permuted blocks (block size 4), stratified by local authority site. The statistician doing the analysis will be blind to the treatment allocation. As the families and practitioners are not blind to treatment allocation (as this would be impossible given the intervention), there is no issue in terms of predictability.

Sample size

Sample size calculations were conducted using R version 4.1.2 (2021-11-01), and the 'pwr' R package.

Parents/carers within the same family will be randomised to the same arm, making this a cluster RCT, given responses within the same family are potentially highly correlated. We might expect, based on our previous research with families and consultation with the TMG (Trial Management Group), an average cluster size within families of up to a maximum of 1.5 parents and a high degree of correlation among parents/carers from the same family, so also allow for an ICC=0.5 (Davé et al., 2008). Following Teerenstra et al. (2012), we also allow for a correlation between baseline and follow-up measures of primary outcome of r=0.5 (this is a conservative estimate based on published SDQ test-retest correlations between 0.74-0.84; Nowak at al.,2008). The sample size is then inflated to account for 20% of families being lost to follow-up at the 12-month post-randomisation follow-up time point.

Allowing for the above assumptions and an effect size to be detected of 0.35 with 90% power and a two-sided alpha of 0.05, a sample size of N=412 (N1 =206, N2=206) is required.

The choice of effect size was based on meta-analytic effects from similar parenting programmes' meta-analyses and on the basis that similar Triple P programmes report large effect sizes that are typically >0.5, some as large as 0.8. On this basis, we reduced to 0.35 given the unique nature of the population in this trial. Further evidence based on individually delivered teen parenting programmes was quite challenging to find well powered RCTs or meta-analyses, so meta-analyses that were slightly outside our age range were also considered.

An efficacy trial for standard teen Triple P report an effect of d=0.62 (Salari et al, 2014). In the other meta-analyses, it was also generally around d=0.6. Given that we are delivering this to a slightly different population, the current planned MDES 0.35 is justifiably conservative (mainly as most studies have been quite small and in a very different population). Considering

the MDES from a clinically meaningful effect size, d=0.35 equates approximately to a 2-point change on our primary outcome, an effect smaller than this is unlikely to provide any meaningful change. With reference to sample size and secondary outcomes, the trial is currently powered on the basis of detecting an appropriate MDES for the primary outcome which is standard practice in all major trials (CONSORT statement; Schulz et al., 2010).

		Protocol	Randomisation
Minimum Detectable Effect Size (MDES)		0.35	
Pre-test/ post-test correlations	level 1 (participant)	0.5	
	level 2 (cluster)	-	
Intracluster correlations (ICCs)	level 1 (participant)	-	
	level 2 (cluster)	0.5	
Alpha[2]		0.05	
Power		0.9	
One-sided or two-sided?		Two-sided	
Average cluster size		1.5	
Number of clusters[3]	intervention	137	
	control	13	
	total	275	

[2] Please adjust as necessary for trials with multiple primary outcomes, 3-arm trials etc. when a Bonferroni correction is used to account for family-wise errors.
[3] Please adjust as necessary e.g., for trials that are randomised at the setting, practitioner or participant level.

		Protocol	Randomisation
Number of participants	intervention	206	
	control	206	
	total	412	

Framework

The trial protocol states that, "Determine whether there is a benefit of support as usual (SAU) plus Standard Teen Triple P (TEEN TRIPLE P) over support as usual (SAU) in improving parent/carer rated adolescent externalising behaviour problems at 6-months post-randomisation in adolescents at the edge of care". Therefore, the trial is on the basis of superiority of the support with additional therapy arm of the trial.

Pilot analysis

Sample size will not be recalculated at any point in the trial, regardless of speed of recruitment. Tables of summary statistics will be produced on all outcomes (primary and secondary).

Characteristics of each trial arm group will be summarised descriptively, both as randomised and as analysed in the primary analysis. Categorical data will be summarised by numbers and percentages. Continuous data that follow a normal distribution will be summarised using means and standard deviations while skewed continuous variables will be summarised using medians and inter-quartile ranges. Histograms and boxplots will be used to check the distribution and possible outliers for continuous variables.

Regarding the sample size for internal pilot, we have pragmatically specified 50% of the overall sample as the recruitment window overall is quite short. We anticipate that recruitment will gather pace so have opted for a longer pilot phase and consequently, a bigger proportion of the final sample. This number does allow us to provide more useful/reliable information from the pilot to assess progression criteria including randomisation and attrition rates. This also follows recommendations from Eldridge et al (2016) that larger pilots for

cluster RCTs are typically more informative and do not follow the same rules of thumb from pilot studies of individually randomised trials.

Internal pilot studies may serve a number of functions and in this case the main functions relate to feasibility of aspects such as recruitment that will not change trial design. Assumptions around sample size can be checked as a part of an internal pilot (Wittes & Brittain, 1990) but we have not proposed to do that given that resources are fixed at the outset of the trial as defined by the funder, YEF.

Timing of final analysis

All outcomes will be analysed following the last 6 month follow up post-randomisation (primary end point) and again for the 12 month follow up post-randomisation (addendum report end point) after the database is locked following the last follow up post randomisation. One month after completion of baseline data collection and data cleaning, the database will be locked to new recruitment and only entry of followup data will be permitted.

Timing of outcome assessment

Outcomes	Data collection timepoints			
	Baseline	**Treatment Phase**	**6 month follow-up**	**12 month follow-up**
Adolescent wellbeing self-report: self-report version of the Strengths and Difficulties Questionnaire	X		X	X
Adolescent wellbeing parent/guardian-report: parent-report version of the Strengths and Difficulties Questionnaire	X		X	X

Parenting Scale Adolescent version (PSA, parent completed)	X		X	X
Dyadic Adjustment Scale (DAS-7)	X		X	X
Kessler 6	X		X	X
Warwick-Edinburgh Mental Well-being Scale (SWEMWBS)	X		X	X
Conflict Behavior Questionnaire (CBQ-20)	X		X	X
Closeness subscale of the Child-Parent Relationship Scale (CPRS, short form)	X		X	X
Family APGAR scale	X		X	X
Parenting Self-Regulation Scale	X		X	X
Out of home placement	X		X	X
Self Report Delinquency Measure	X			X

Statistical Principles

Levels of confidence and p-values

All confidence intervals presented will be 95% and two-sided. In addition, applicable statistical tests for the primary analysis will be 2-sided and will be performed using a 5% significance level.

Adjustment of multiplicity

The overall type I error rate for testing support as usual (SAU) plus Standard Teen Triple P (TEEN TRIPLE P) trial arm over the control arm SAU only for the primary endpoint will be controlled at the 2-sided 0.05 significance level (i.e. no correction required as we have a single primary outcome). Secondary analyses will control the family-wise error rate using the Holm method. Secondary outcomes will be analysed and corrected in the following order:

(1) Parent reported - Adolescent behavioural and emotional problems
(2) Parenting practices
(3) Parent and Adolescent reported prosocial behaviours
(4) Adolescent reported peer relationships
(5) Interparental outcome
(6) Parent mental health
(7) Parent Wellbeing
(8) Conflict Behavior
(9) Child-parent relationship
(10) Family functioning
(11) Parent self-regulation – the Parenting Self-Regulation
(12) Delinquency

The Holm method, in a stepwise way, computes the significance levels depending on the P value based rank of hypotheses. For the ith ordered hypothesis $H(i)$, the specifically adjusted significance level is computed:

$$\alpha'(i) = \frac{\alpha}{m - i + 1}$$

where m is the number of hypothesis tests.

The observed P value $p(i)$ of hypothesis $H(i)$ is then compared with its corresponding $\alpha'(i)$ for statistical inference; and each hypothesis will be tested in order from the smallest to largest P values $(H(1), \ldots, H(m))$ The comparison will immediately stop when the first $p(i) \geq \alpha'(i)$ is observed $(i = 1, \ldots, m)$ and hence all remaining hypotheses of $H(j)$ $(j = i, \ldots, m)$ are directly declared non-significant without requiring individual comparison.

Adherence and protocol deviations

Definition and assessment of adherence

TEEN TRIPLE P attendance/engagement data will be recorded in logs by practitioners, including: start date of Parent/carer engagement with the intervention and number of sessions completed.

The number of sessions delivered will be recorded by practitioners in Session Summary forms and any implementation challenges recorded.

Adherence: session attendance (green=≥ 75% of families attend at least the first 8 of the 10 sessions; amber= 50-74%; red=<50%). Note: This is inline with progression criteria (page 17-18 of the trial protocol) and in line with the required minimum number of specific sessions attended (core set) as advised by the intervention developers.

Presentation of adherence

The number and % of participants for green, amber and red thresholds for proportion of scheduled sessions attended will be presented in a table for i) randomisation to 6 months follow up, and ii) 6 months to 12 months follow up. Results will be provided by treatment group including adherence to SAU sessions.

Definition of protocol deviation

Any deviation from the randomised intervention plan as detailed in the protocol will be considered as a protocol deviation.

Presentation of protocol deviation

Prospective, planned deviations or waivers to the protocol will not be allowed, e.g. participants who do not meet the eligibility criteria or restrictions specified in the trial protocol will not be enrolled.

Any accidental protocol deviations will be adequately documented on the relevant forms and reported to the Chief Investigators immediately.

Deviations from the protocol which occur frequently will be addressed immediately and if appropriate will be classified as a serious breach.

The final analysis will also present the proportions of protocol deviations in a table (following intention to treat principle).

Analysis population

Parents and foster carers identified by case-holding social workers and edge of care teams across six local authorities (Cambridgeshire, Peterborough, Birmingham, Gloucestershire, London Borough Merton, and Wirral) and their young people aged 11-15 years and at the edge of care. The intention-to-treat population will include all randomised patients, , according to the treatment they were randomised to receive.

Study population

Screening data

The following summaries will be presented for all screened Parents/ carers and Young people (overall and by local authority):

Enrolment: the number of days recruiting, the number of Parents/carers/CYP screened, the number of Parents/carers/CYP recruited, the number of screened Parents/carers/CYP not recruited, and the reason for non-recruitment.

This information will be included into the CONSORT flow diagram (see appendix for template).

Eligibility

Families are eligible for the trial if they meet all the following inclusion criteria and none of the exclusion criteria apply. All queries about family eligibility should be directed to the Trial Manager before randomisation/recruitment.

Definition of *Edge of Care*

Edge of Care refers to children/young people who either:

- have not entered into care as they have been assessed and the LA has chosen to support them and their families through alternative provisions/services. Or
- they are being considered for care but have not entered into local authority care.

Inclusion criteria

- Families of young people aged 11-15 years (inclusive) determined as being on the *edge of care*

Exclusion criteria

- Families where one or more parent has received a multi-session parenting programme covering similar content to Triple P over the previous 12 months
- Families where one or more parent is currently receiving a multi-session parenting programme covering similar content to Triple P or any multi-component manualised family intervention, such as Multi-Systemic Therapy

The number of ineligible participants randomised, if any, will be reported, with reasons for ineligibility. Ineligible participants will be removed from the data and not included into the analysis.

Recruitment

A CONSORT flow diagram (appendix A) will be used to summarise the number of CYP who were:

- assessed for eligibility at screening
- eligible at screening
- ineligible at screening*
- eligible, consented and randomised
- eligible but not randomised*
- received the randomised allocation
- did not receive the randomised allocation*
- lost to follow-up*
- discontinued the intervention*
- randomised and included in the primary analysis
- randomised and excluded from the primary analysis*

*reasons will be provided.

Withdrawal/Follow up

Level of withdrawal

The participants service as usual will not be affected at any time by declining to participate or withdrawing from the trial because they will still receive services as usual. If a participant initially consents but subsequently withdraws from the trial, clear distinction will be made as to what aspect of the trial the participant is withdrawing from. These aspects will be:

- Withdrawal from intervention (TEEN TRIPLE P only)
- Partial withdrawal from future follow-up data collection (e.g., some questionnaires, interviews)
- Withdrawal from previously collected data, prior to data analysis

Participants cannot withdraw from the trial but still receive the intervention, if they withdraw from the trial then they will receive usual services only. All participants will be included in the primary analysis unless they withdraw their consent for the use of their data.

Timing of withdrawal

Participants have the right to withdraw consent for participation in any aspect of the trial at any time.

Reasons for withdrawal

Participants who consent and subsequently withdraw should complete the trial withdrawal form or the withdrawal form should be completed on the participant's behalf by the site staff/ trial team based on information provided by the participant. This withdrawal form should be sent to the Trial email address. Any queries relating to potential withdrawal of a participant should be forwarded to the Trial Manager.

Presentation of withdrawal/Loss to follow up

The number and % of participants that have withdrawn/loss to follow up from the study will be presented in a table for all stages. Results will be provided by treatment group.

Baseline participant characteristics

List of baseline data

Participants will be screened at site and eligibility will be assessed. Potential participant details will be passed from the trial site to the trial team in Warwick. The trial team will contact the participant as per their preferred choice of data collection to take consent and complete the baseline data:

PARENT/CARER:

- DOB (month and year only)
- Sex/gender
- Ethnicity
- If English is their first language
- Questions identifying of young person has a special educational needs or mental health condition:

  a. "*Does your child /young person have any physical or mental health conditions or illnesses lasting or expected to last 12 months or more? "*
  b. "If yes, *Does your child / young person's condition or illness / do any of their conditions or illnesses reduce their ability to carry-out day-to-day activities?*"
  c. "*Does your child/young person have special educational needs? "*

d. "If 'yes', what are the reason for his/her special education needs / additional support needs?"

- Baseline outcome measures completed

    1. CYP wellbeing parent/ guardian-report: parent-report version of the Strengths and Difficulties Questionnaire
    2. Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999) parent report internalising problems scale, adolescent report externalising and internalising problems scales at 6 and 12 months.
    3. Parenting practices: Parenting Scale Adolescent version (PSA) parent report
    4. Adolescent prosocial behaviours – The Prosocial behaviour subscale of the SDQ (Goodman, 1999). Adolescents and parents report at 6 and 12 months.
    5. Adolescent peer relationships – The Peer Relationship Problem subscale of the SDQ (Goodman, 1999)
    6. Interparental outcome - Dyadic Adjustment Scale (DAS-7) is a 7-item measure (Sharpley & Cross 182; Hunsley et al., 2001)
    7. Parent mental health – the Kessler 6 (Kessler et al., 2003
    8. Parent Wellbeing - The Warwick-Edinburgh Mental Well-being Scale (SWEMWBS; Tennant et al., 2007
    9. Conflict Behavior Questionnaire (CBQ-20) (Robin & Foster, 1989)
    10. Child-parent relationship – the Closeness subscale of the Child-Parent Relationship Scale (CPRS, short form) measures (Pianta, 1995)
    11. Family functioning – the Family APGAR scale (Adaptability, Partnership, Growth, Affection and Resolve; APGAR; Smilkstein, 1978).
    12. Parent self-regulation – the Parenting Self-Regulation Scale is a 12-item parent-completed measure of parental-regulation (Tellegen et al., 2022).
    13. Out of home placement

YOUNG PERSON:

o DOB (month and year only)

o Sex/gender

o Who they live with and any changes in living arrangements between baseline and follow-up, and if they are being looked after

o Whether they are in school

o Type of school

o School year

o Ethnicity

o If English is their first language

o GP contact details

• Baseline outcome measures completed (WASI-II is to be completed with researcher assistance [telephone, teleconferencing, or face-to-face])

1. Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999) self report internalising problems scale, adolescent report externalising and internalising problems scales at 6 and 12 months.
2. Conflict Behavior Questionnaire (CBQ-20) (Robin & Foster, 1989)
3. Delinquency - Self Report Delinquency Measure (SRDM; Smith & McVie, 2003)

Descriptive statistics

Characteristics of each trial arm group will be summarised descriptively, both as randomised and as analysed in the primary analysis.

Categorical data will be summarised by numbers and percentages. Continuous data that follow a normal distribution will be summarised using means and standard deviations while skewed continuous variables will be summarised using medians and inter-quartile ranges. Histograms and boxplots will be used to check the distribution and possible outliers for continuous variables. Tests of statistical significance will not be undertaken for baseline characteristics (Senn, 1994); rather the clinical importance of any imbalance will be noted.

Analysis

Outcome definitions

Primary outcome(s)

The proposed primary outcome measure for this trial is the parent/guardian version of the Strengths and Difficulties Questionnaire (SDQ) for externalising behaviours and will be used to assess CYP well-being. The SDQ is a robust and well-validated measure of behavioural and emotional problems (Deighton et al., 2014); measured over the preceding 6 months.

Timing, units, and derivation of primary

The primary outcome is collected at baseline, 6 months, and 12-months post-randomisation. The SDQ externalising score is a derived score using the hyperactivity and conduct sub-

domains, following Deighton et al. (2014) and the units are a relative measure of externalising behaviour.

Baseline and 6-month follow up data will only be used in the primary analysis. The 12 month follow up data will be used both in the addendum primary analysis and in an additional follow up analysis including all three time points in a linear mixed model (see section 6.4 – longitudinal follow up analyses). The SDQ consists of 25 items which are each scored on a 3-point Likert scale (0, 1, 2). Externalising scores - Ranges from 0-20 and is generated by summing the scores of the conduct and hyperactivity subscales.

List of secondary outcomes

Secondary outcome measures include:

- Emotional and behavioural difficulties and prosocial behaviour: the Parent report and adolescent self-report versions of the Strengths and Difficulties Questionnaire (SDQ) will be used (including internalising, externalising, peer relationship, and prosocial behaviours). The SDQ is a robust and well-validated measure of behavioural and emotional problems (Deighton et al., 2014); measured over the preceding 6 months.
- Parenting practices: Parenting Scale Adolescent version (PSA, parent completed) is a short form of the Parenting Scale (Irvine et al., 1999) which assesses dysfunctional discipline practices in parents. It is an adaptation of the Parenting Scale (Arnold, et al., 1993) for parents of adolescents
- Interparental outcome - Dyadic Adjustment Scale (DAS-7) is a 7-item measure (Sharpley & Cross 182; Hunsley et al., 2001) assessing the relationship quality of couples. The DAS-7 assesses relationship satisfaction and the degree to which the couple agrees on matters of importance to the relationship.
- Parent mental health – the Kessler Psychological Distress Scale (Kessler 6) is a six-item screening tool for serious mental illness in the general population (Kessler et al., 2003). It will be completed by parents.
- Parent Wellbeing - The Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) is a measure of mental wellbeing (Tennant et al., 2007). The short (7 item) version will be completed by parents.
- Conflict Behavior Questionnaire (CBQ-20) (Robin & Foster, 1989) assesses adolescent-parent communication, conflict and relations. Both the adolescent and parent report versions will be completed.
- Child-parent relationship – the Closeness subscale of the Child-Parent Relationship Scale (CPRS, short form) measures (Pianta, 1995) will be completed by parents.

- Family functioning – the Family APGAR scale (Adaptability, Partnership, Growth, Affection and Resolve; APGAR) measures satisfaction with family functioning (Smilkstein, 1978). This 5-item measure will be completed by parents.
- Parent self-regulation – the Parenting Self-Regulation Scale is a 12-item parent-completed measure of parental-regulation (Tellegen et al., 2022).
- Delinquency – the Self Report Delinquency Measure (Smith & McVie, 2003) is a 15 item measure of antisocial behaviours (e.g., burglary, violence) and will be completed by adolescents at 12 month follow up.

Order of testing

Secondary outcomes are tested in the following order:

(1) Parent reported - Adolescent behavioural and emotional problems
(2) Parenting practices
(3) Parent and Adolescent reported prosocial behaviours
(4) Adolescent reported peer relationships
(5) Interparental outcome
(6) Parent mental health
(7) Parent Wellbeing
(8) Conflict Behavior
(9) Child-parent relationship
(10) Family functioning
(11) Parent self-regulation – the Parenting Self-Regulation
(12) Delinquency

Timing, units and derivation of secondaries

Secondary outcomes are collected at baseline, 6-months and 12-months post-randomisation.

- Emotional and behavioural difficulties: Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. The SDQ consists of 25 items which are each scored on a 3-point Likert scale (0, 1, 2). Three subscales will be used: i) Externalising - Ranges from 0-20 and is generated by summing the scores of the conduct and hyperactivity subscales; ii) internalising - Ranges from 0-20 and is generated by summing the emotional and peer problems subscales; iii) peer relationships – ranges from 0-10 and is generated by summing peer relationships items; and iv) prosocial behaviours – ranges from 0-10 and is generated by summing prosocial behaviour items.
- Gang Affiliation: Baseline, 6-month, and 12-month follow up data will be collected but only baseline and 6-month follow up will be used in secondary outcome analysis. The

primary analysis will be repeated in the 12 month followup addendum report. There are 15 binary (yes/no) items that are summed giving a range 0-15 total score. A total score of 7 or more would indicate risk of gang affiliation and would suggest early intervention support is provided.

- Parenting scale Adolescent: Baseline, 6-month and 12 month follow up data will be collected. Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. The PSA is a 30 item measure with each item scored on a 7 point scale. A total score is created by summing all items, with low scores indicating good parenting and high scores indicating dysfunctional parenting.
- Dyadic Adjustment Scale (DAS-7): Baseline, 6-month and 12 month follow up data will be collected. Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. The DAS-7 is comprised of varying response scales, including both ordinal and Likert scales.
    - Items 1–3 use a 6-point ordinal scale (from 5 = 'Always Agree' to 0 = 'Always Disagree').
    - Items 4–6 also use a 6-point ordinal scale (from 0 = 'Never' to 5 = 'More Often').
    - Item 7 is rated on a 7-point Likert scale (from 0 = 'Extremely Unhappy' to 6 = 'Perfect')

  The total score for the DAS-7 is the sum of the responses to the seven items. The resultant score ranges from 0 to 36, with a higher score indicating more positive relationship quality. Scores less than 21 are considered to indicate a relationship in distress.

- Parent mental health – the Kessler 6 (K6) : Baseline, 6-month and 12 month follow up data will be collected. Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. The K6 is scored using a 5-level response scale, ranging from 0 to 4 (0 = none of the time, 1 = a little of the time, 2 = some of the time, 3 = most of the time, 4 = all of the time); this generates a scoring scale with a range of 0 to 24.
- Parent Wellbeing - The Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) - short version: Baseline, 6-month and 12 month follow up data will be collected. Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. The SWEMWBS is scored by first summing the scores for each of the seven items, which are scored from 1 to 5. The total raw scores are then transformed into metric scores using the SWEMWBS conversion table which can be found here: https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/howto/swemwbs_raw_score_to_metric_score_conversion_table.pdf

- Conflict Behavior Questionnaire (CBQ-20): Baseline, 6-month and 12 month follow up data will be collected. Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. This is a 20 item scale with each item scored true or false and totalled score between 0-20. High scores represent more negative communications.
- Closeness subscale of the Child-Parent Relationship Scale (CPRS, short form) - Baseline, 6-month and 12 month follow up data will be collected. Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. The seven items are scores on a 5-point scale. Scores for the closeness subscale range between 7-35. A higher score on the closeness questions suggest the parent/child relationship is characterised by warmth, affection and open communication. Therefore, a higher score is desirable on the closeness subscale.
- the Family APGAR scale: Baseline, 6-month and 12 month follow up data will be collected. Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. Items are scored as follows: 'Almost always' (2 points), 'Some of the time' (1) point, or 'Hardly ever' (0). The scores for each of the five questions are then totalled. A score of 7 to 10 suggests a highly functional family. A score of 4 to 6 suggests a moderately dysfunctional family. A score of 0 to 3 suggests a severely dysfunctional family.
- Parent self-regulation – the Parenting Self-Regulation Scale: Baseline, 6-month and 12 month follow up data will be collected. Baseline and 6-month follow up data will only be used in the main secondary analysis, but will be repeated in the 12 month followup addendum report. Items are scored on a 7-point Likert-type scale ranging from (1) “strongly disagree” to (7) “strongly agree”. A total score is calculated by summing all 12 items, with higher scores reflecting higher levels of parent self-regulation.
- Delinquency – the Self Report Delinquency Measure: Baseline and 12 month follow up data will be collected. Baseline and 12-month follow up data will only be used in the 12 month followup addendum report. The SRDM is a derived total score following Smith & McVie (2003) and the units are a relative measure of delinquency. The SDRM is a measure comprising 15-items pertaining to antisocial behaviours (e.g., burglary, violence). It requires CYP to respond with yes or no with reference to a time-period (6 months). They then report the estimated frequency of the behaviour, and whether they have ever been caught. Each items frequency is scored 0-5, 6-10 is scored 6 and 11+ is scored 11. Minimum score would be 0 and maximum number of delinquent behaviours would be 165 (15x11). On this basis, we are likely to have a skewed continuous distribution, so some transformation may be required after inspection of model residuals. In addition, there may be a number of individuals where this is their first time in a custody unit, so there is a possibility of floor effects depending on the

frequency of their delinquent behaviour. A higher number of delinquent behaviours is bad, so a reduction in the outcome indicates an effective treatment.

Analysis methods

List of methods and presentation

Internal pilot study

Statistical analysis for internal pilot feasibility outcomes will be primarily descriptive. Feasibility outcomes (primary outcome measures and all secondary measures) will be estimated as frequencies and percentages, means and standard deviations, or medians and interquartile ranges as appropriate. Feasibility outcomes will be assessed against the pre-specified progression criteria.

Primary analysis (6 month)

Our primary analysis will include all randomised participants who provide outcome data (i.e., a intention to treat basis) and compare mean scores between arms on the SDQ externalising behaviour at 6-months post-randomisation using a linear mixed model, adjusting for baseline SDQ score and local authority. A random intercept will also be included to account for family level clustering at level 2 (individual at level 1 and family at level 2).

$$\boldsymbol{L1}: Y_{ik} = \beta_{0k} + \beta_{1k} SDQ\_BL_{ik} + \beta_{2k} local\_authority_{ik} + r_{ik}, \quad r_{ik} \sim N(0, \sigma^2)$$

$$\boldsymbol{L2}: \beta_{0k} = \gamma_{00} + \gamma_{20} TX_k + \mu_{0k}, \qquad \mu_{0k} \sim N(0, \tau_2^2)$$

$$\beta_{1k} = \gamma_{10}$$

where, Y_{ik} are the SDQ scores; SDQ_BL_{ik} are the baseline SDQ scores; TX_{ik} is the treatment/control variable indicator; $Local\_authority_{ik}$ are indicator of local authority (strata, 6 groups: Cambridge, Peterborough, Wirral, Birmingham, London Borough Merton, Gloucestershire); μ_{0k} is the random intercept term for family and r_{ik} is the individual level variation.

We will use simple coding for the contrast of local authority, so that our intercept retains the grand mean and nominally use Birmingham as our reference level.

Distributional assumptions for the primary linear model will be checked and alternative methods are listed in section “assumption checking”.

If insufficient random effect variance is found, we will reduce the model to a linear regression with cluster robust standard errors (Green & Vavreck, 2008). The alternative primary analysis would be specified as follows:

$$Y_i = \beta_0 + \beta_1 SDQ\_BL_i + \beta_2 local\_authority_i + \beta_3 TX_i + r_i,\ r_i \sim N(0, \sigma^2\ )$$

Secondary outcomes analysis

Secondary outcomes will be analysed following the same method as the primary outcome. However, the distributions of these secondary outcomes will be assessed prior to conducting the analysis, if range restriction is present or a binary outcome (out of home placement), we will consider using a generalized linear mixed model (GLMM) instead (Poisson or logistic respectively). All measures have scores that are integer numbers and positive scores, so the Poisson distribution under this condition is completely acceptable if needed. The GLMM is denoted as follows:

$$\boldsymbol{L1}: g(Y_{ik}) = \beta_{0k} + \beta_{1k} SDQ\_BL_{ik} + \beta_{2k} local\_authority_{ik} + r_{ik},\ r_{ik} \sim N(0, \sigma^2\ )$$

$$\boldsymbol{L2}: \beta_{0k} = \gamma_{00} + \gamma_{02} TX_k + \mu_{0k}, \qquad \mu_{0k} \sim N(0, \tau_2^2)$$

$$\beta_{1k} = \gamma_{10}$$

Note: $g(.) = log_e(.)$, where $g(.)$ Is the log link function for the secondary outcome measures for example, if a count model was more appropriate for the scores, whereas the primary outcome, $g(Y_i) = Y_i$.

For continuous outcome measures, the effect size and 95% confidence interval will be calculated using given in Hedges (2007) for cluster randomised designed analysed via multilevel models and allowing for unequal cluster sizes. According to the two-level LMM for primary outcome given that we have individuals nested within family, a sample estimate of the effect size equivalent to Hedges' g with 95% confidence interval is defined as:

$$\hat{\Delta}_g = \frac{\widehat{\beta_1}}{S_T} \sqrt{1 - \frac{2(n-1)\rho}{N-2}}$$

Where $\widehat{\beta_1}$ is the adjusted mean difference in SDQ externalising score between trial arms; S_T is the within group pooled standard deviation

$S_T^2 = \frac{\sum_{i=1}^{m^I} \sum_{j=1}^{n_i^I} (Y_{ij}^I - Y_{i..}^I)^2 + \sum_{i=1}^{m^C} \sum_{j=1}^{n_i^C} (Y_{ij}^C - Y_{i..}^C)^2}{N-2}$ Where 'm^I' is the total number of families in the intervention sample, and 'n^I' the total number of parent/carers (equivalent definitions apply for the control group, but with the 'C' designation). $Y_{i..}^I$ and $Y_{i..}^C$ are the mean outcomes among intervention and control families respectively.

The remaining part of the $\hat{\Delta}_g$ equation makes the adjustment for clustering. The two intra - class correlation coefficients at the family (ρ) level are defined as follows,

$$\rho = \frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2} = \frac{\sigma_B^2}{\sigma_T^2},$$

where σ_B^2 is the between-family variance, and σ_W^2 is the within-family variance

If outcomes are categorical, counts or binary, the effect sizes will be reported as odds ratios or incident rate ratios. All parameter estimates from the models will be reported with 95% confidence intervals.

We will report the ICC from the two-level model which is defined as:

$$\boldsymbol{ICC} = \frac{\boldsymbol{\sigma^2}_{\boldsymbol{family}}}{\boldsymbol{\sigma^2+\sigma^2}_{\boldsymbol{family}}},$$

where σ^2 is the residual variance, and $\sigma^2{}_{family}$ is the random intercept variance according to family (family level clustering).

Covariate adjustment

We will assess any imbalance of baseline covariates for possible inclusion in the model where large imbalances are noted. However, due to the sample size, we do not anticipate substantial issues in this respect.

Assumption checking

1. Linearity – plotting residuals vs predictor(s). If a structure is present, then transformation or an alternate model specification is required (i.e. GLMM).

2. Homogeneity of variance – variance of the residuals across groups is the same. There is scope to fit models allowing for heterogeneous groups, but the setup is different (Generalized linear mixed model - GLMM).

3. Residuals are approximately normally distributed – plotting QQ plot

Alternative methods if distributional assumptions not met

If distributional assumptions are not satisfied, as appropriate, a generalized linear mixed model with alternate link function will be used. Alternatively, data transformation could be used but use of the GLMM is preferable.

Sensitivity analyses

• Exploring the impact of missing data on trial outcomes by investigating likely missing data mechanisms and re-fitting the primary outcome within a multiple imputation framework (including exploring MAR and MNAR mechanisms via delta-based controlled multiple imputation);

• Adding practitioner as a source of clustering in the model. The primary analysis will be rerun including a third level where individual parent will be nested in family and families will be nested by practitioner. An additional random intercept will be included to account for practitioner level clustering at level 3 as follows:

$$\boldsymbol{L1}: Y_{ijk} = \beta_{0jk} + \beta_{1jk} SDQ\_BL_{ijk} + \beta_{2jk} local\_authority_{ijk} + r_{ijk}, \; r_{ijk} \sim N(0, \sigma^2)$$

$$\boldsymbol{L2}: \beta_{0jk} = \gamma_{00k} + \gamma_{20k} TX_k + \mu_{0jk}, \qquad \mu_{0jk} \sim N(0, \tau_{00}^2)$$

$$\beta_{1jk} = \gamma_{10k}$$

$$\boldsymbol{L3}: \gamma_{00k} = \delta_{000} + U_{0k}, \qquad U_{0k} \sim N(0, \varphi_{00}^2)$$

$$\gamma_{10k} = \delta_{100}$$

where, Y_{ijk} are the SDQ scores; SDQ_BL_{ijk} are the baseline SDQ scores; TX_{ijk} is the treatment/control variable indicator; $Local\_authority_{ijk}$ are indicator of local authority (strata, 6 groups: Cambridge, Peterborough, Wirral, Birmingham, London Borough Merton, Gloucestershire); μ_{0jk} is the random intercept term for family, U_{0k} is the random intercept term for practitioner, and r_{ik} is the individual level variation.

• Exploring the impact of different levels of intervention receipt on outcomes. We will use two-stage least squares instrumental variables regression to examine the effect of the intervention in those who receive varying levels of it.

Fidelity: The fidelity score is the average proportion of sessions content delivered. Fidelity items will be scored 0, 0.5, and 1. Total fidelity session score will be dependent on the session and will range between 20 and 30. This is based on agreed progression criteria and Triple P guidance on minimum number of sessions required in their opinion as the intervention developer.

Adherence: Number of sessions attended (Progression criteria: green=≥ 75% of families attend at least the first 8 of the 10 sessions; amber= 50-74%; red=<50%). The proportion of sessions attended out of a maximum of ten will be the instrumental variable in this analysis.

For both fidelity and adherence, a Two Stage Least Square approach will be used to estimate the model and Huber-White standard errors reported which are robust to clustering. The R packages 'ivpack' and 'ivreg' will be used to implement the two-stage instrumental variable analysis (Jiang & Small, 2014; Fox Kleiber, & Zeileis, 2021). Compliance (session adherence) will be instrumented by the intervention allocation (Angrist & Imbens, 1995). However, robust SE will be reported instead of including random effects in to the model. The stage 1 model is defined as follows:

$$Compliance_k = \beta_0 + \beta_1 TX_k + \varepsilon_{jk}$$

Predicted values for, $Compliance_k$, from the stage 1 model will be included in the stage 2 model, as follows:

$$Y_{ik} = \beta_0 + \beta_1 \widehat{compliance}_k + \beta_2 baseline_{ik} + \beta_3 Local\_authority_k + \beta_4 \widehat{compliance}_k \cdot Local\_authority_k + r_{ik}$$

Subgroup analyses

In addition to the primary and secondary outcomes, we have considered that the following outcomes may moderate the primary outcome of this trial.

- Ethnicity,
- In-care vs family home status: all instances of out of home placement will be recorded at each follow-up time point. This will include the reason for and duration of out of home placement.
- Learning disabilities (LD): Children and young people will be invited to complete two subtests of the Wechsler Abbreviated Scale of Intelligence-II (WASI-II; Wechsler, 2011) to index their Verbal IQ. This scale is to be administered with a researcher (face-to-face, telephone, videoconferencing). The two subsets are to be included are Vocabulary and Similarities.

A moderation analysis will adjust the primary analysis with the inclusion of the moderator as a main effect and interaction between moderator and randomised group indicator. For example, the LD moderator analysis is as follows:

$$\boldsymbol{L1}: Y_{ik} = \beta_{0k} + \beta_{1k} SDQ\_BL_{ik} + \beta_{2k} local\_authority_{ik} + \beta_{3k} LD_{ik} + r_{ik}, \quad r_{ik} \sim N(0, \sigma^2)$$

$$\boldsymbol{L2}: \beta_{0k} = \gamma_{00} + \gamma_{02} TX_k + \mu_{0k}, \qquad \mu_{0k} \sim N(0, \tau_2^2)$$

$$\beta_{1k} = \gamma_{10}$$

$$\beta_{2k} = \gamma_{20}$$

$$\beta_{3k} = \gamma_{30} + \gamma_{20} TX_k$$

where, Y_{ik} are the SDQ scores; SDQ_BL_{ik} are the baseline SDQ scores; TX_{ik} is the treatment/control variable indicator; $Local\_authority_{ik}$ are indicator of local authority (strata); LD_{ik} is learning disability status; μ_{0k} is the random intercept term for family; and r_{ik} is the individual level variation.

In addition to interpreting the magnitude and statistical significance of interactions, plots of the interactions will also be examined. These analyses will be hypothesis generating in nature only (i.e., will not be confirmatory and only indicate whether further research targeting the intervention may be warranted).

Missing data

Exploring the impact of missing data on the primary trial outcome by investigating likely missing data mechanisms and re-fitting the primary outcome within a multiple imputation framework (including exploring MAR and MNAR mechanisms via delta-based controlled multiple imputation).

We will summarise the extent of missing data in all outcomes and their respective control variables. A full multiple imputation strategy will be used if more than 5% of data in the primary model is missing. Alternatively, we will impute if more than 10% of data for a single variable is missing. We will use the multiple imputation by chained equations approach via the mice package in R (Van Buuren and Groothuis-Oudshoorn, 2011) and generate 10 imputed datasets. We will then estimate the intervention effect for each imputed dataset and pool the results using Rubin's combination rules for standard errors.

Missing data in item level data

The sum scores will not be imputed directly. Any missing item level data will be imputed using the chained equation approach. Each item's imputation model will use other items and covariates specified in the primary analysis model as predictors.

Following creation of the imputed datasets, the corresponding total scores will be calculated using the imputed item level data. All imputed datasets will then fit the primary and secondary models and pool estimates following Rubin's rules (Rubin, 2004; van Buuren, 2018).

Primary outcome

Given that each item is an ordered category, we will use an ordinal regression within the imputation model for each item.

Secondary outcome Similarly, the correct link function will be used according to the item's structure for each of the secondary outcomes, i.e. binary or ordered categorical accordingly. Therefore, a logistic or ordinal model will be used in the imputation for these items.

Additional analyses

Longitudinal followup analyses (12 month only) – Primary outcome

We will fit linear mixed models, accounting for repeated post-randomisation measures (6- and 12-months post-randomisation) within participants, adjusting for local authority to investigate the overall effect of the intervention on post-randomisation measures.

$$\boldsymbol{L1}: Y_{ijk} = \beta_{0jk} + \beta_{1jk} time_{ijk} + r_{ijk},\ r_{ijk} \sim N\left(0, \sigma^2{}_{|X}\right)$$

$$\boldsymbol{L2}: \beta_{0jk} = \gamma_{00k} + \mu_{0jk}, \qquad \mu_{0jk} \sim N(0, \tau_2^2)$$

$$\beta_{1jk} = \gamma_{100} + \mu_{1jk}, \qquad \mu_{1jk} \sim N(0, \tau_2^2)$$

$$\boldsymbol{L3}: \gamma_{00k} = \delta_{000} + \delta_{001} local\_authority_k + \delta_{001} TX_k + \varsigma_{0k},\ \varsigma_{0k} \sim N(0, \tau_3^2)$$

Longitudinal followup analyses (12 month only) - Secondary outcomes

Secondary outcomes will be analysed similarly, but

$$\boldsymbol{L1}: g(Y_{ijk}) = \beta_{0jk} + \beta_{1jk} time_{ijk} + r_{ijk},\ r_{ijk} \sim N\left(0, \sigma^2{}_{|X}\right)$$

$$\boldsymbol{L2}: \beta_{0jk} = \gamma_{00k} + \mu_{0jk}, \qquad \mu_{0jk} \sim N(0, \tau_2^2)$$

$$\beta_{1jk} = \gamma_{100} + \mu_{1jk}, \qquad \mu_{1jk} \sim N(0, \tau_2^2)$$

$$\boldsymbol{L3}: \gamma_{00k} = \delta_{000} + \delta_{001} local\_authority_k + \delta_{001} TX_k + \varsigma_{0k},\ \varsigma_{0k} \sim N(0, \tau_3^2)$$

For the remaining secondary outcomes, their effect sizes will be reported as either Hedges' g (same as primary outcome) or rate ratios (other secondary outcomes, exponentiated parameter estimates), given that generalized linear mixed effects models with log link function are used to model the data and that the measures are positively scored integers with some amount of skew anticipated (Barnett and Dobson, 2008).

Mediation analyses (12 months report only)

Exploratory mediation analyses may also be carried out to examine variables at 6 months that may mediate intervention effects between baseline and 12-month follow-up. Mediation analyses will use the same two-level model setup as in the main analysis, but will have the following additional steps:

- Initial model, outcome regressed on all independent variables except mediator (same as main analysis).

$$Y_i = \beta_0 + \beta_1 SDQ_{BL_i} + \beta_2 local\_authority_i + \beta_3 TX_i + \varepsilon_i,$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

- Second, the same model is fitted but the dependent measure is exchanged for the mediator.

$$M_i = \beta_0 + \beta_1 SDQ_{BL_i} + \beta_2 local\_authority_i + \beta_3 TX_i + \varepsilon_i,$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

- Finally, the first model is refitted but with the inclusion of the mediator as an additional independent variable.

$$Y_i = \beta_0 + \beta_1 SDQ_{BL_i} + \beta_2 local\_authority_i + \beta_3 TX_i + \beta_4 M_i + \varepsilon_i,$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

Using these model parameter estimates and the 'mediate' function from the R package 'mediation', the average causal mediation effects can be estimated following the procedure from Imai et al. (2010). The mediation analyses will use the bootstrapped procedure as this method has no distributional assumptions on the indirect effect and is generally more robust (Bollen & Stine, 1990).

Harms

The number (and percentage) of patients experiencing each AE/SAE will be presented for each treatment arm categorised by severity. For each patient, only the maximum severity experienced of each type of AE will be displayed. The number (and percentage) of occurrences of each AE/SAE will also be presented for each treatment arm. No formal statistical testing will be undertaken.

Statistical software

All statistical analyses will use R version 4.1.2 (2021-11-01) with additional packages: tidyverse, pwr, VGAM, lme4, lmerTest, performance, mice, psych, ivreg, and ivpack.

References

Joshua D. Angrist & Guido W. Imbens (1995) Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity, Journal of the American Statistical Association, 90:430, 431-442, 10.1080/01621459.1995.10476535

Barnett, A.G., & Dobson, A.J. (2008). An Introduction to Generalized Linear Models (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315182780

Bell, B. A., Ferron, J. M., & Kromrey, J. D. (2008). Cluster size in multilevel models: the impact of sparse data structures on point and interval estimates in two-level models. JSM Proceedings, Section on Survey Research Methods, 1122-1129.

Bollen, K.A. and Stine, R. (1990). Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability. Sociological Methodology. Vol. 20:115-. DOI: 10.2307/271084

Borm G.F., Fransen J., Lemmens W.A. (2007). A simple sample size formula for analysis of covariance in randomized clinical trials. J Clin Epidemiol, 60(12):1234-8. doi: 10.1016/j.jclinepi.2007.02.006.

Coulman, E., Hastings, R., Gore, N. et al. The Early Positive Approaches to Support (E-PAtS) study: study protocol for a feasibility cluster randomised controlled trial of a group programme (E-PAtS) for family caregivers of young children with intellectual disability. Pilot Feasibility Stud 6, 147 (2020). https://doi.org/10.1186/s40814-020-00689-9

Davé, S., Nazareth, I., Senior, R. et al. A Comparison of Father and Mother Report of Child Behaviour on the Strengths and Difficulties Questionnaire. Child Psychiatry Hum Dev 39, 399–413 (2008). https://doi.org/10.1007/s10578-008-0097-6

Eldridge SM, Costelloe CE, Kahan BC, Lancaster GA, Kerry SM. How big should the pilot study for my cluster randomised trial be? Statistical Methods in Medical Research. 2016;25(3):1039-1056. doi:10.1177/0962280215588242

Fox, F., Kleiber, C. and Zeileis, A. (2021). ivreg: Instrumental-Variables Regression by '2SLS', '2SM', or '2SMM', with Diagnostics. R package version 0.6-1. https://CRAN.R-project.org/package=ivreg

Green, D., & Vavreck, L. (2008). Analysis of Cluster-Randomized Experiments: A Comparison of Alternative Estimation Approaches. Political Analysis, 16(2), 138-152. doi:10.1093/pan/mpm025

Hedges, L. V. (2007). Effect Sizes in Cluster-Randomized Designs. Journal of Educational and Behavioral Statistics, 32(4), 341–370. https://doi.org/10.3102/1076998606298043

Imai, K., Keele, L., Tingley, D. (2010). A general approach to causal mediation analysis. *Psychol Methods,* 15(4), p309-34. doi: 10.1037/a0020761. PMID: 20954780.

Jiang, Y. and Small, D. (2014). ivpack: Instrumental Variable Estimation. R package version 1.2. https://CRAN.R-project.org/package=ivpack

Nowak, C., Heinrichs, N. A (2008). Comprehensive Meta-Analysis of Triple P-Positive Parenting Program Using Hierarchical Linear Modeling: Effectiveness and Moderating Variables. Clin Child Fam Psychol Rev, 11, 114–144.

Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys (Vol. 81). John Wiley & Sons.

Senn, S. (1994), Testing for baseline balance in clinical trials. Statist. Med., 13: 1715-1726. https://doi.org/10.1002/sim.4780131703

Smith, D.J., & McVie, S. (2003). Theory and method in the Edinburgh study of youth transitions and crime. British Journal of Criminology, 43(1), 169-195.

Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. DOI: 10.18637/jss.v045.i03.

Teerenstra, S., Eldridge, S., Graff, M.J., Hoop, E., & Borm, G.F. (2012). A simple sample size formula for analysis of covariance in cluster randomized trials. *Statistics in Medicine, 31*.

van Buuren, S. (2018). Flexible Imputation of Missing Data, Second Edition (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259

Wechsler D. Wechsler Abbreviated Scale of Intelligence - Second Edition. Bloomington, United States: Pearson; 2011.

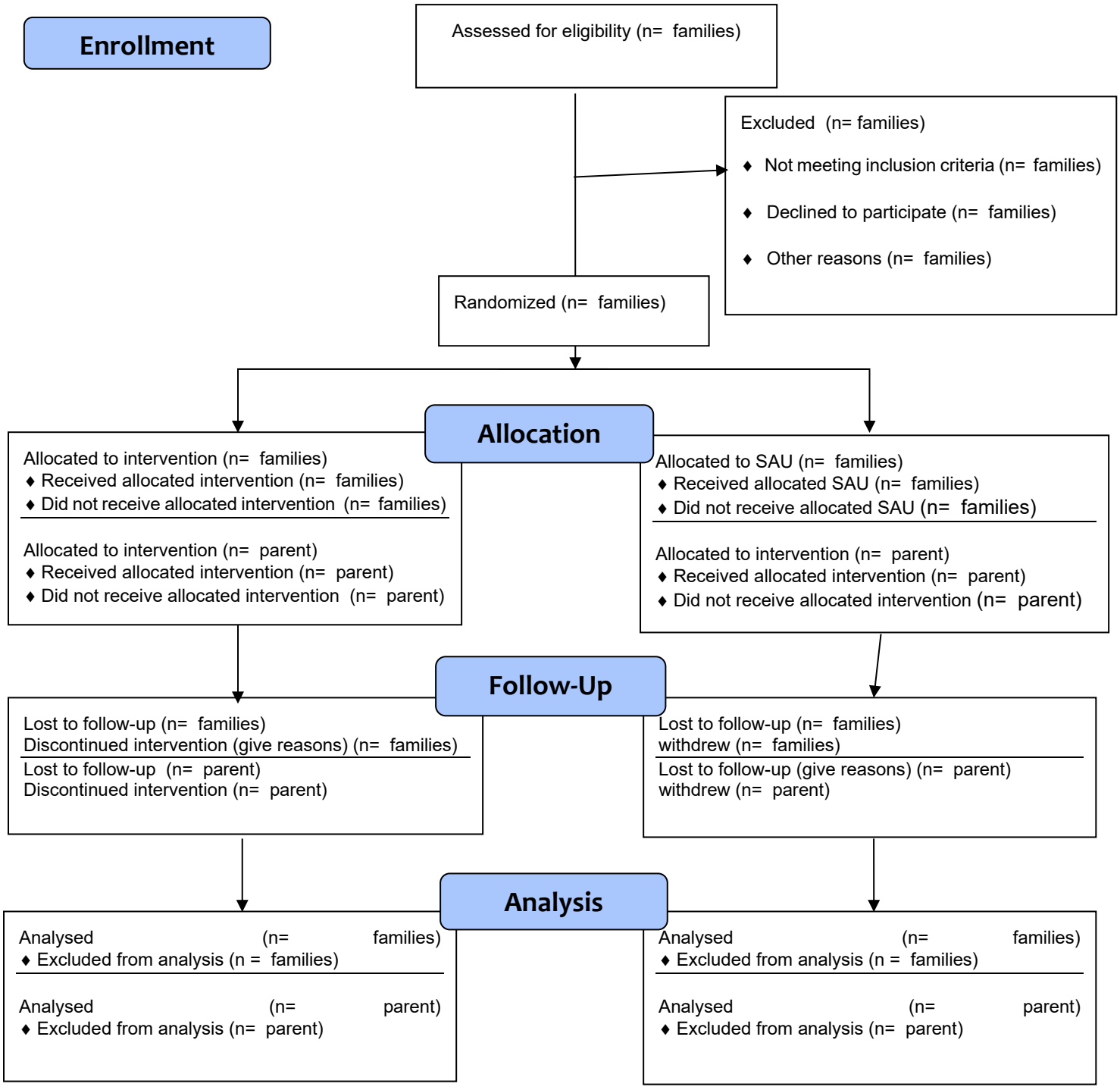
Wittes, J. and Brittain, E. The role of internal pilot studies in increasing the efficiency of clinical trials. Statist. Med. 1990. 9(1-2): doi:10.1002/sim.4780090113

SAP Deviation Log

Document number:		**Document version:**	
Reason for deviation:			

Appendix A

CONSORT 2010 Flow Diagram for Cluster RCT



YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413